

The Karcher Mean of Points on SO_n

Knut Hüper

joint work with

Jonathan Manton (Univ. Melbourne)

Knut.Hueper@nicta.com.au

National ICT Australia Ltd.



Contents



- **Introduction**



Contents



- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**
- **Karcher mean on SO_n**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**
- **Karcher mean on SO_n**
- **Cost function, gradient and Hessian**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**
- **Karcher mean on SO_n**
- **Cost function, gradient and Hessian**
- **Newton-type algorithm**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**
- **Karcher mean on SO_n**
- **Cost function, gradient and Hessian**
- **Newton-type algorithm**
- **Convergence results**

- **Introduction**
 - **Centroids, Karcher mean, Fréchet mean**
 - **The Euclidean case**
- **Motivation, why SO_n**
- **Radii of convexity and injectivity**
- **Karcher mean on SO_n**
- **Cost function, gradient and Hessian**
- **Newton-type algorithm**
- **Convergence results**
- **Discussion, outlook**



Several ways to define a centroid x_C



Given $x_1, \dots, x_k \in \mathbb{R}^n$.



Several ways to define a centroid x_C



Given $x_1, \dots, x_k \in \mathbb{R}^n$.

1) As the sum

$$x_c := \frac{1}{n} \sum_{i=1}^k x_i.$$



Several ways to define a centroid x_C



Given $x_1, \dots, x_k \in \mathbb{R}^n$.

1) As the sum

$$x_c := \frac{1}{n} \sum_{i=1}^k x_i.$$

2) Equivalently, to ask the vector sum

$$\overrightarrow{xx_1} + \dots + \overrightarrow{xx_k}$$

to vanish.



Several ways to define a centroid x_C



3) (Appolonius of Perga) As unique minimum of

$$x_c := \operatorname{argmin}_{x \in \mathbb{R}} \sum_{i=1}^k \|x - x_i\|^2.$$



Several ways to define a centroid x_C



3) (Appolonius of Perga) As unique minimum of

$$x_c := \operatorname{argmin}_{x \in \mathbb{R}} \sum_{i=1}^k \|x - x_i\|^2.$$

4) More generally, assign to each x_i a mass m_i , $\sum m_i = 1$. By induction

$$x_{c_{1,2}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}$$
$$x_{c_{1,2,3}} = \frac{(m_1 + m_2) x_{c_{1,2}} + m_3 x_3}{m_1 + m_2 + m_3}, \dots$$

Also works on spheres.



Several ways to define a centroid x_C



5) Axiomatically:

Let $\Phi : \mathbb{R}^n \times \dots \times \mathbb{R}^n \supset \Xi \rightarrow \mathbb{R}^n$ be a rule mapping points to its centroid.



Several ways to define a centroid x_C



5) Axiomatically:

Let $\Phi : \mathbb{R}^n \times \dots \times \mathbb{R}^n \supset \Xi \rightarrow \mathbb{R}^n$ be a rule mapping points to its centroid.

Axioms:

(A1) Φ is symmetric in its arguments.

(A2) Φ is smooth.

(A3) Φ commutes with the induced action of SE_n on $\mathbb{R}^n \times \dots \times \mathbb{R}^n$.

(A4) If $\Omega \subset \mathbb{R}^n$ is an open convex ball then Φ maps $\Omega \times \dots \times \Omega$ into Ω .

- (A1) Centroid is independent of the ordering of the points.**
- (A2) Small changes in the location of the points causes only small changes in x_c .**
- (A3) Invariance w.r.t. translation and rotation.**
- (A4) Centroid lies in the "same region" as the points themselves. Especially, $\Phi(x, \dots, x) = x$.**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**
 - **smoothing data**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**
 - **smoothing data**
 - **plate tectonics**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**
 - **smoothing data**
 - **plate tectonics**
 - **sequence dep. continuum modeling of DNA**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**
 - **smoothing data**
 - **plate tectonics**
 - **sequence dep. continuum modeling of DNA**
 - **comparison theorems (diff. geometry)**



Why centroids on manifolds?



- **Engineering, Mathematics, Physics**
 - **statistical inferences on manifolds**
 - **pose estimation in vision and robotics**
 - **shape analysis and shape tracking**
 - **fuzzy control on manifolds (defuzzification)**
 - **smoothing data**
 - **plate tectonics**
 - **sequence dep. continuum modeling of DNA**
 - **comparison theorems (diff. geometry)**
 - **stochastic flows of mass distributions on manifolds (jets in gravitational field)**



The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$



The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$

Facts:



The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$

Facts:

a) SO_n is a Lie group,



The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$

Facts:

- a) SO_n is a Lie group,
- b) is in general not diffeomorphic to a sphere,



The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$

Facts:

- a) SO_n is a Lie group,
- b) is in general not diffeomorphic to a sphere,
- c) can be equipped with a Riemannian metric,
therefore notion of distance is available,



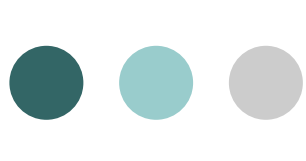
The special orthogonal group

$$SO_n$$


$$SO_n := \{X \in \mathbb{R}^{n \times n} \mid X^T X = I, \det X = 1\}.$$

Facts:

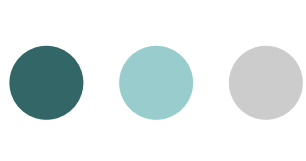
- a) SO_n is a Lie group,
- b) is in general not diffeomorphic to a sphere,
- c) can be equipped with a Riemannian metric, therefore notion of distance is available,
- d) is compact and connected, but in general not simply connected.



Geometry of SO_n



a) We think of SO_n as a submanifold of $\mathbb{R}^{n \times n}$.



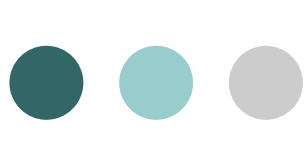
Geometry of SO_n



a) We think of SO_n as a submanifold of $\mathbb{R}^{n \times n}$.

b) Tangent space

$$T_X SO_n \cong \{XA \mid A \in \mathbb{R}^{n \times n}, A^T = -A\}.$$



Geometry of SO_n



a) We think of SO_n as a submanifold of $\mathbb{R}^{n \times n}$.

b) Tangent space

$$T_X SO_n \cong \{XA \mid A \in \mathbb{R}^{n \times n}, A^\top = -A\}.$$

c) (Scaled) Frobenius inner product on $\mathbb{R}^{n \times n}$

$$\langle U, V \rangle = \frac{1}{2} \operatorname{tr}(V^\top U)$$

restricts to

$$\langle XU, XV \rangle = \frac{1}{2} \operatorname{tr}(V^\top U), \quad U, V \in T_X SO_n.$$

Gives Riemannian metric on SO_n .



Geometry of SO_n cont'd



d) Let $X \in SO_n$, $\Omega^T = -\Omega \in \mathbb{R}^{n \times n}$.



Geometry of SO_n cont'd



d) Let $X \in SO_n$, $\Omega^\top = -\Omega \in \mathbb{R}^{n \times n}$.

$$\begin{aligned}\gamma : \mathbb{R} &\rightarrow SO_n, \\ t &\mapsto X \cdot e^{t \cdot \Omega}\end{aligned}$$

is a geodesic through $X = \gamma(0)$.



Geometry of SO_n cont'd



d) Let $X \in SO_n$, $\Omega^\top = -\Omega \in \mathbb{R}^{n \times n}$.

$$\begin{aligned}\gamma : \mathbb{R} &\rightarrow SO_n, \\ t &\mapsto X \cdot e^{t \cdot \Omega}\end{aligned}$$

is a geodesic through $X = \gamma(0)$.

$$\int_0^T \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle^{\frac{1}{2}} dt$$

is minimal (for T not too large..)



Geometry of SO_n cont'd



e) Squared distance between any two points

$$X, Y \in SO_n$$



Geometry of SO_n cont'd



e) Squared distance between any two points

$$X, Y \in SO_n$$

$$\begin{aligned} d^2(X, Y) &= \frac{1}{2} \min_{\substack{A^\top = -A \\ \exp(A) = X^\top Y}} \text{tr}(AA^\top) \\ &= -\frac{1}{2} \text{tr}(\log(X^\top Y))^2 \end{aligned}$$



Centroid of SO_n by axioms



Let

$$\Xi \subset SO_n \times \cdots \times SO_n$$

be open and consider $\Phi : \Xi \rightarrow SO_n$.



Centroid of SO_n by axioms



Let

$$\Xi \subset SO_n \times \cdots \times SO_n$$

be open and consider $\Phi : \Xi \rightarrow SO_n$.

(A1) Φ is symmetric in its arguments.



Centroid of SO_n by axioms



Let

$$\Xi \subset SO_n \times \cdots \times SO_n$$

be open and consider $\Phi : \Xi \rightarrow SO_n$.

(A1) Φ is symmetric in its arguments.

(A2) Φ is smooth.



Centroid of SO_n by axioms



Let

$$\Xi \subset SO_n \times \cdots \times SO_n$$

be open and consider $\Phi : \Xi \rightarrow SO_n$.

(A1) Φ is symmetric in its arguments.

(A2) Φ is smooth.

(A3) Φ commutes with left and right translation.



Centroid of SO_n by axioms



Let

$$\Xi \subset SO_n \times \cdots \times SO_n$$

be open and consider $\Phi : \Xi \rightarrow SO_n$.

(A1) Φ is symmetric in its arguments.

(A2) Φ is smooth.

(A3) Φ commutes with left and right translation.

(A4) If $\Omega \subset SO_n$ is an open convex ball then Φ maps $\Omega \times \cdots \times \Omega$ into Ω .



Notion of convexity



$\Omega \subset SO_n$ is defined to be convex if for any $X, Y \in SO_n$ there is a unique geodesic wholly contained in Ω connecting X to Y and such that it is also the unique minimising geodesic in SO_n connecting X to Y .



Notion of convexity



$\Omega \subset SO_n$ is defined to be convex if for any $X, Y \in SO_n$ there is a unique geodesic wholly contained in Ω connecting X to Y and such that it is also the unique minimising geodesic in SO_n connecting X to Y .

A function $f : \Omega \rightarrow \mathbb{R}$ is convex if for any geodesic $\gamma : [0, 1] \rightarrow \Omega$, the function $f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ is convex in the usual sense, that is,

$$f(\gamma(t)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(1)), \quad t \in [0, 1].$$



Notion of convexity cont'd



Maximal convex ball (centered at the identity I_n)



Notion of convexity cont'd



Maximal convex ball (centered at the identity I_n)

$$B(I, r) = \{X \in SO_n \mid d(I, X) < r\}.$$

r_{conv} is the largest r s.t. $B(I, r)$ is convex and $d(I, X)$ is convex on $B(I, r)$.



Notion of convexity cont'd



Maximal convex ball (centered at the identity I_n)

$$B(I, r) = \{X \in SO_n \mid d(I, X) < r\}.$$

r_{conv} is the largest r s.t. $B(I, r)$ is convex and $d(I, X)$ is convex on $B(I, r)$.

Theorem: For SO_n it holds $r_{\text{conv}} = \frac{\pi}{2}$.



Injectivity radius



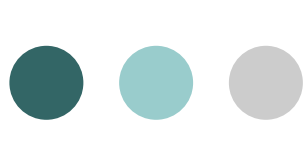
For $\mathfrak{so}_n := \{A \in \mathbb{R}^{n \times n} \mid A^\top = -A\}$ **let**

$$\exp : \mathfrak{so}_n \rightarrow SO_n,$$

$$\Psi \mapsto \exp(\Psi),$$

and

$$B(0, \rho) = \{A \in \mathfrak{so}_n \mid \frac{1}{2} \operatorname{tr} A^\top A < \rho^2\}.$$



Injectivity radius



For $\mathfrak{so}_n := \{A \in \mathbb{R}^{n \times n} \mid A^\top = -A\}$ let

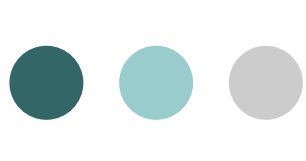
$$\exp : \mathfrak{so}_n \rightarrow SO_n,$$

$$\Psi \mapsto \exp(\Psi),$$

and

$$B(0, \rho) = \{A \in \mathfrak{so}_n \mid \frac{1}{2} \operatorname{tr} A^\top A < \rho^2\}.$$

The injectivity radius r_{inj} of \mathfrak{so}_n is the largest ρ s.t. $\exp|_{B(0, \rho)}$ is a diffeomorphism onto its image.



Injectivity radius



For $\mathfrak{so}_n := \{A \in \mathbb{R}^{n \times n} \mid A^\top = -A\}$ let

$$\exp : \mathfrak{so}_n \rightarrow SO_n,$$

$$\Psi \mapsto \exp(\Psi),$$

and

$$B(0, \rho) = \{A \in \mathfrak{so}_n \mid \frac{1}{2} \operatorname{tr} A^\top A < \rho^2\}.$$

The injectivity radius r_{inj} of \mathfrak{so}_n is the largest ρ s.t. $\exp|_{B(0, \rho)}$ is a diffeomorphism onto its image.

Theorem: For \mathfrak{so}_n it holds $r_{\text{inj}} = \pi$.



Karcher mean on SO_n



Let $\Omega \subset SO_n$ be open.



Karcher mean on SO_n



Let $\Omega \subset SO_n$ be open.

A Karcher mean of $Q_1, \dots, Q_k \in SO_n$ is defined to be a minimiser of

$$f : \Omega \rightarrow \mathbb{R},$$

$$f(X) = \sum_{i=1}^k d^2(Q_i, X).$$



Karcher mean on SO_n



Let $\Omega \subset SO_n$ be open.

A Karcher mean of $Q_1, \dots, Q_k \in SO_n$ is defined to be a minimiser of

$$f : \Omega \rightarrow \mathbb{R},$$

$$f(X) = \sum_{i=1}^k d^2(Q_i, X).$$

Existence, uniqueness?

Theorem (MH'04):
The critical points of

$$f : \Omega \rightarrow \mathbb{R},$$
$$f(X) = \sum_{i=1}^k d^2(Q_i, X)$$

are precisely the solutions of

$$\sum_{i=1}^k \log(Q_i^\top X) = 0.$$

Theorem (MH'04):

The Karcher mean is well defined and satisfies axioms (A1)-(A4) of a centroid on the open set

$$\Xi = \bigcup_{Y \in SO_n} B(Y, \pi/2) \times \cdots \times B(Y, \pi/2).$$

Theorem (MH'04):

The Hessian of f represented along geodesics

$$\left. \frac{d^2}{dt^2} (f \circ \gamma)(t) \right|_{t=0}$$

is always positive definite.



f , $\text{grad } f$ and Hessian explicitly



$$f : \Omega \rightarrow \mathbb{R},$$

$$f(X) = \sum_{i=1}^k d^2(Q_i, X) = - \sum_{i=1}^k \frac{1}{2} \text{tr}(\log(X^\top Q_i))^2.$$



f , $\text{grad } f$ and Hessian explicitly



$$f : \Omega \rightarrow \mathbb{R},$$

$$f(X) = \sum_{i=1}^k d^2(Q_i, X) = - \sum_{i=1}^k \frac{1}{2} \text{tr}(\log(X^\top Q_i))^2.$$

$$\begin{aligned} D f(X) X A &= - \sum_{i=1}^k \text{tr}(\log(Q_i^\top X) A) \\ &= \underbrace{\left\langle 2X \sum_{i=1}^k \log(Q_i^\top X), X A \right\rangle}_{= \text{grad } f(X)}. \end{aligned}$$



f , $\text{grad } f$ and **Hessian explicitly**



$$\frac{d^2}{d\varepsilon^2} f \left(X e^{\varepsilon A} \right)_{\varepsilon=0} = \text{vec}^\top A \cdot \mathcal{H}(X) \cdot \text{vec } A$$



f , $\text{grad } f$ and Hessian explicitly



$$\frac{d^2}{d\varepsilon^2} f \left(X e^{\varepsilon A} \right)_{\varepsilon=0} = \text{vec}^\top A \cdot \mathcal{H}(X) \cdot \text{vec } A$$

with $(n^2 \times n^2)$ -matrix

$$\mathcal{H}(X) := \sum_{i=1}^k Z_i(X) \coth(Z_i(X))$$



f , $\text{grad } f$ and Hessian explicitly



$$\frac{d^2}{d\varepsilon^2} f \left(X e^{\varepsilon A} \right)_{\varepsilon=0} = \text{vec}^\top A \cdot \mathcal{H}(X) \cdot \text{vec } A$$

with $(n^2 \times n^2)$ –**matrix**

$$\mathcal{H}(X) := \sum_{i=1}^k Z_i(X) \coth(Z_i(X))$$

and

$$Z_i(X) := \frac{I_n \otimes \log(Q_i^\top X) + \log(Q_i^\top X) \otimes I_n}{2}$$

Given $Q_1, \dots, Q_k \in SO_n$, compute a local minimum of f .

Step 1: Set $X \in SO_n$ to an initial estimate.

Step 2: Compute $\sum_{i=1}^k \log(Q_i^\top X)$.

Step 3: Stop if $\|\sum_{i=1}^k \log(Q_i^\top X)\|$ is suff. small.

Step 4: Compute the update direction

$$\text{vec } A_{\text{opt}} = -(\mathcal{H}(X))^{-1} \sum_{i=1}^k \text{vec}(\log(Q_i^\top X))$$

Step 5: Set $X := X e^{A_{\text{opt}}}$.

Step 6: Go to Step 2.



Results



Theorem (MH'04):

The algorithm is an intrinsic Newton method.

Theorem (MH'04):

The algorithm is an intrinsic Newton method.

Theorem:

If the algorithm converges, then it converges locally quadratically fast.



Discussion, outlook



- **Need simple test to ensure that update step in algorithm remains in open convex ball \Rightarrow global convergence.**



Discussion, outlook



- **Need simple test to ensure that update step in algorithm remains in open convex ball \Rightarrow global convergence.**
- **Different RM, e.g. Cayley-like, gives different function, geodesics, etc., but typically**
$$\|KM_{\text{cay}} - KM_{\text{exp}}\| \ll 1.$$



Discussion, outlook



- **Need simple test to ensure that update step in algorithm remains in open convex ball \Rightarrow global convergence.**
- **Different RM, e.g. Cayley-like, gives different function, geodesics, etc., but typically**
$$\|KM_{\text{cay}} - KM_{\text{exp}}\| \ll 1.$$
- **$(\mathcal{H}(X))^{-1}$ via EVD.**



Discussion, outlook



- **Need simple test to ensure that update step in algorithm remains in open convex ball \Rightarrow global convergence.**
- **Different RM, e.g. Cayley-like, gives different function, geodesics, etc., but typically $\|KM_{\text{cay}} - KM_{\text{exp}}\| \ll 1$.**
- **$(\mathcal{H}(X))^{-1}$ via EVD.**
- **Quasi-Newton (rank-one updates).**



Discussion, outlook



- **Linear convergent algorithm
(joint work with Robert Orsi, ANU)**

$$X_{i+1} = X_i e^{\frac{1}{k} \sum_{j=1}^k \log(X_i^\top Q_j)}$$



Discussion, outlook



- **Linear convergent algorithm
(joint work with Robert Orsi, ANU)**

$$X_{i+1} = X_i e^{\frac{1}{k} \sum_{j=1}^k \log(X_i^\top Q_j)}$$

- **Centroids on homogeneous (symmetric) spaces.**



Discussion, outlook



- **Linear convergent algorithm
(joint work with Robert Orsi, ANU)**

$$X_{i+1} = X_i e^{\frac{1}{k} \sum_{j=1}^k \log(X_i^\top Q_j)}$$

- **Centroids on homogeneous (symmetric) spaces.**
- **Project with NICTA vision/robotic program (Richard Hartley) to treat SE_3 case.**